Optimization Method for Trajectory Data Based on Satellite Doppler Velocimetry

Junzhuo Li1*, Wenyong Li2, Guan Lian3

School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China1, 2
Guilin University of Electronic Technology Guangxi Key Laboratory of Intelligent Transportation, Guilin 541004, China3

Abstract—Due to cost and energy consumption limitations, there are significant differences in the positioning capabilities of mobile terminals, resulting in unsatisfactory quality of trajectory data. In this paper, satellite Doppler data is used to optimize trajectory data. First, the system state equation is established by the kinematic relationship between the measured velocity and position, and the static linear Kalman filter estimates the optimal system state. Then a dynamic Kalman filter system is established by correlating the measurement error matrix parameters of the Kalman filter with the vertical dilution of precision of satellite positioning. Finally, the whole-day trajectory of a taxi in Shenzhen was visualized, and the deviation between the trajectory points and the urban road was calculated to compare the optimized and non-optimized taxi trajectories. The results show that the proposed optimization method can effectively reduce the deviation between trajectory points and urban roads, and this method can be used to process vehicle trajectory data in urban traffic research.

Keywords—Urban transportation; Kalman Filter; information fusion; trajectory data

I. INTRODUCTION

The rise of mobile Internet and location-based services (LBS) has generated a large amount of trajectory data. These data provide abundant materials for analyzing and researching social behavior and economic activities and have significant scientific research value. In the study of social phenomena, trajectory data have many advantages [1, 2], such as: 1) Trajectory data is an objective record and description of the position changes of mobile terminals and users, which is not affected by subjective feelings and has objectivity. 2) Mobile Internet has the characteristics of timeliness, which makes trajectory data update very fast and can provide timely feedback on social phenomenon and behavior changes. 3) With the popularity of mobile terminals such as mobile phones, location-based services have been widely applied, generating a massive amount of trajectory data. 4) Mobile terminals with positioning functions, such as mobile phones, tablets, taxis, and shared bicycles, can generate various trajectory data through various methods such as GPS, cellular network positioning, WI-FI positioning, etc.

During the collection process of trajectory data, there are inevitably systematic errors due to positioning technology. For example, the positioning accuracy of GSP is related to factors such as the signal reception ability of the receiver, calculation methods, local shading conditions, atmospheric conditions, and the number of observable satellites, and the positioning accuracy of cellular networks is closely related to network format, solution methods, and the density of signal base stations. Due to cost, energy consumption, safety, and other reasons, mobile terminals such as mobile phones can usually only receive L1 carrier signals and C/A codes and cannot use differential positioning based on carrier phase. Therefore, optimizing their trajectories is very crucial [3].

Improving positioning and information communication technology is a direct way to improve positioning accuracy. For example, the third-generation GPS satellite launched in 2018 can launch a new civilian signal L1C in the 1575.42MHz band, making GPS more compatible with other GNSS systems and enabling receivers to receive more satellite signals [4]. In addition, the third-generation GPS satellite is equipped with a better accurate rubidium atomic clock and has higher signal transmission power [5, 6]. With the upgrading of communication technology, the accuracy of positioning technology based on cellular networks is also unceasingly improving [7]. In 3GPP Release 16, a new Down Link-Position Reference (DL-PRS) is brought in, and various positioning technologies with better performance, such as DL-TDOA, UL-TDOA, DL-AOD, UL-AOA, E-CID, could be used in the fifth-generation communication network [8]. In the fifth-generation communication network, the dense network makes TDOA and DOA positioning more reliable, Massive Multiple Input Multiple Output (Massive MIMO) technology provides an Infrastructure for AOA positioning, and lower network latency improves the accuracy of time-based positioning methods [9-11].

In addition to improvements in positioning and communication technology, data processing and fusion can also improve the quality of trajectory data. Using spatial clustering and filtering algorithms can make trajectory data smoother, avoiding deviations or anomalies in trajectory data [12-15]. By integrating information beyond positioning data, such as electronic maps, Doppler velocimetry, and vehicle inertial measurement units, the quality of trajectory data can also be improved [16-18].

In this study, the trajectory data is optimized using the velocity measured by GPS satellites based on the Doppler effect. The structure of this article is as follows: In Section II, the system state equation is established according to the kinematics principle, and then the acquisition of the measurement value of the system state equation is introduced, that is, the pseudo-range positioning method and the Doppler velocity measurement method of the satellite. In Section III, the static Kalman filter is used to estimate the optimal system state. Then the dynamic Kalman filter is used to estimate the
optimal system state by correlating the system measurement matrix with the satellite positioning Dilution of Precision (DOP). In Section IV, the daily trajectory of a taxi in Shenzhen was visualized, and the deviation between the trajectory and the urban road network was calculated to compare the optimized and non-optimized taxi trajectories.

II. ESTABLISH SYSTEM STATE EQUATION

The state equation of the Kalman filter system mainly includes two prediction and measurement processes. According to the laws of kinematics, the system state equation can be obtained as Formula (1). In this formula, $o_k$ is the system state at time $k$, which includes $(x^s, y^s)$ meaning the position of the object and $(v_x^s, v_y^s)$ meaning the speed of the object. $A_k$ is the system state transition matrix, $w_k$ is the system noise, which conforms to the Gaussian distribution with the mean value of 0, and the variance of $Q$, $v_k$ is the measured noise, which conforms to the normal distribution with the mean value of 0 and the variance of $R$.

$$
\begin{bmatrix}
o_t \\
z_t
\end{bmatrix} = 
\begin{bmatrix}
o_k \\
v_k
\end{bmatrix} +
\begin{bmatrix}
x^s & \Delta t & x^s \\
y^s & \Delta t & y^s \\
v_x^s & 1 & v_x^s \\
v_y^s & 1 & v_y^s
\end{bmatrix} 
+ N - (0, Q)
$$

$$
z_{t} = o_{t} + N - (0, R)
$$

A. System Measurement: GPS Pseudo-Range Positioning

GPS has the advantages of fast positioning speed, accurate positioning, low cost, and wide coverage. The positioning results can be obtained by solving Formula (2) with four unknown variables $x$, $y$, $z$, $\delta$. When the receiver can receive four satellite signals, the equation has a unique solution; when the received signal is less than 4, the GPS satellite cannot independently calculate the receiver position; When the received signals are more than 4, the least square method is usually used to solve the overdetermined equations. When many satellite signals can be received, the more the number of equations, the more stable the solution result, and the more minor effect of one satellite measurement deviation on the positioning result, making the positioning result more accurate and stable [19, 20]. The positioning accuracy of GPS is closely related to the number of satellite signals received. Due to cost and energy consumption limitations, mobile terminals’ satellite signal reception capacity is commonly insufficient, and high-rise buildings could block GPS signal transmission in urban areas. Many reasons will affect the reception of satellite signals, resulting in inaccurate positioning, so improving the accuracy of trajectory data is crucial [21].

$$(x' - x)^2 + (y' - y)^2 + (z' - z)^2 = (\rho' - \delta c)^2$$

In Formula (2), $[x^s, y^s, z^s]$ is the coordinates of the received satellite; $[x, y, z]$ is the coordinates of the receiver; $\delta$ is the receiver’s clock error; $c$ is the speed of light; $\rho'$ is the pseudo-range of the received satellite, which can be calculated by Formula (3).

$$\rho' = (r' + \delta - \delta' + I + T)c + \epsilon'$$

In Formula (3), $r'$ is the time required for the signal to be transmission between the receiver and the satellite under vacuum conditions, $\delta$ is the clock error of the satellite, $I$ is the ionospheric delay, $T$ is the tropospheric delay, and $\epsilon'$ is the error of the pseudo-range.

B. System Measurement: Satellite Doppler Velocimetry

Formula (3) calculates the derivative of time to obtain Formula (4).

$$\rho' = (r' + \delta - \delta' + I + T)c + \epsilon'$$

In Formula (4), $\rho'$ is the rate of change of pseudo-range, which can be calculated by Doppler frequency shift [22], as in Formula (5); $\dot{r}'c$ is the change rate of the geometric distance between the receiver and satellite, as shown in Formula (6); $\delta$ and $\delta'$ are frequency drift of receiver and satellite; $I$ and $T$ are the change rate of ionospheric delay and tropospheric delay, which are small enough to be ignored [23].

$$\rho' = -\lambda (f - f')$$

In Formula (5), $\lambda$ is the wavelength of the transmitted signal, $f'$ and $f$ are the frequency of the signal received by the receiver and the frequency of the signal transmitted by satellite, respectively.

$$\dot{r}'c = (v' - v)I'$$

In Formula (6), $v' = [v'_x, v'_y, v'_z]^T$ is the travel speed of the satellite; $v = [v_x, v_y, v_z]^T$ is the travel speed of the terminal device or user, which is an unknown variable to be solved; $I'$ is the directional unit vector of the receiver, which can be calculated according to Formula (7).

$$I' = \frac{1}{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}$$

Substitute Formula (5-6) into Formula (4) to get Formula (8), and then sort out all unknown variables into the left side to get Formula (9) with four unknown variables $v_x, v_y, v_z, \delta$. When the number of satellite signals that can be received is greater than or equal to four, the travel speed of the mobile terminal or user can be obtained by solving Formula (9).

$$-\lambda (f - f') = (v' - v)I' + \delta - \delta' + \epsilon'$$
\[ vI' - \delta = \lambda f - \lambda f' + v'I - \delta' + \delta' \tag{9} \]

### III. Kalman Filtering System

Kalman filter is an optimal estimation method using the system state equation. When the measurement variance is known, the Kalman filter can estimate the optimal state of the system from a series of measurements containing noise [24, 25]. This Section introduces the basic calculation process of the Kalman filter, and the static Kalman filter method is used to calculate the changing trend of system Kalman gain and system state uncertainty. Then, the random error of the measured value of the system is correlated with the DOP of satellite positioning to establish a dynamic Kalman filter system.

#### C. Static Kalman Filter System

The linear Kalman filter method calculation includes two stages of prediction and update involving five formulas.

In the prediction stage, the calculation of the linear Kalman filter includes:

- **System state prediction equation.** It uses the previous state \( \hat{\delta}_{k-1} \) to infer the current state \( \hat{\delta}_k \), such as Formula (10). In Formula (10), \( A \) is the state transition matrix, which establishes the relationship between the current system state and the system state at the previous moment through kinematics laws.

\[ -\lambda(f - f') = (v' - v)I' + \hat{\delta} - \hat{\delta}' + \hat{\delta}' \tag{10} \]

- **Prediction equation of system uncertainty.** It calculates the prior estimate \( \hat{P}_k \) of the current system state uncertainty according to the uncertainty \( \hat{P}_{k-1} \) of the previous moment and the system’s random error \( Q \) of the observation value, as shown in Formula (11).

\[ -\lambda(f - f') = (v' - v)I' + \hat{\delta} - \hat{\delta}' + \hat{\delta}' \tag{11} \]

In the update stage, the calculation of the linear Kalman filter includes:

\[ K_k = \frac{P_k}{P_k + R} \tag{12} \]

- **System state update equation.** It calculates the current system optimal state \( \hat{\delta}_k \) according to the prior estimate \( \hat{\delta}_k \) of the system state, the current measured value \( z_k \), and the Kalman gain \( K_k \), such as Formula (13).

\[ \hat{\delta}_k = \hat{\delta}_k + K_k(z_k - \hat{\delta}_k) \tag{13} \]

- **Update equation of system uncertainty.** It calculates the optimal uncertainty \( P_k \) of the current system state according to the prior estimated uncertainty \( P_k \) and Kalman gain \( K_k \), such as Formula (14). In Formula (14), \( E \) is an identity matrix.

\[ P_k = (E - K_k)P_k \tag{14} \]

The calculation flow of the whole Kalman filter is shown in Fig. 1. The upper level of Fig. 1 includes two initial and measurement modules responsible for data input. The initialization module inputs the iterative initial values of \( \hat{\delta}_0 \) and \( P_0 \), and the parameters needed for random error \( R \), systematic error \( Q \), and transition matrix \( A \) to establish the system equation. The measurement module continuously inputs the measured values \( z_k \) in the iterative calculation process of the system. The middle level of Fig. 1 is the core of the calculation in the Kalman filter, which uses Formula (10-14) to iterate the system’s state, including the prediction and update modules. The lower level of Fig. 1 is responsible for the output of the system’s optimal state and uncertainty.
This study uses the Doppler velocity measurement and pseudo-range positioning results of GPS to establish the system state, with a calculation step of $\Delta t = 10$, and the settings of the system error and random error of measurement are shown in Formula (15-16).

\[
Q = \begin{bmatrix}
0.5 & 0.1 & 0.1 & 0.5 \\
0.1 & 0.5 & 0.1 & 0.5 \\
0.1 & 0.1 & 0.5 & 0.1 \\
0.5 & 0.5 & 0.1 & 0.5
\end{bmatrix}
\]

\[
R = \begin{bmatrix}
2.5 & 0.5 & 0.5 & 2.5 \\
0.5 & 0.5 & 0.5 & 0.5
\end{bmatrix}
\]

According to the established system state equation, the variation trend of the Kalman gain and uncertainty with iteration is calculated and plotting Fig. 2. This Kalman filter system is static because the parameters $A$, $Q$, and $R$ of the system are constant. According to Fig. 2, the Kalman gain and uncertainty gradually decrease and converge with the calculation iteration in the static system.

D. Dynamic Kalman Filtering System

The transition matrix $A$ is invariant, determined by the kinematic relationship between position and speed in the system state equation, as well as the system error matrix $Q$ will not be easily changed in specific equipment. However, the number and geometric distribution of received satellites in space change over time according to the ephemeris and can affect the measurement random error $R$. When $R(t)$ is added to the system instead of $R$, a dynamic Kalman filter system can be obtained [26]. Generally, GPS measurement random error can be indicated by the DOP. Fig. 3 shows the DOP and the number of received satellites in the study area on the day of the experiment.

Fig. 3 shows five different DOPs, geometric DOP, time DOP, position DOP, horizontal DOP, and vertical DOP, among which the horizontal DOP is the most suitable for establishing a functional relationship with $R(t)$. At 9:50, the maximum value of horizontal DOP was 1.66, and at 00:00, the minimum value of horizontal DOP was 0.73. Therefore, the variation trend of the Kalman gain and system state uncertainty with iteration was calculated when $R(t) = 1.66$ and $R(t) = 0.73$, as shown in Fig. 4. In the dynamic Kalman filter system, the Kalman gain and the system state uncertainty will vary within the values of the filling area in Fig. 4. With the iterative calculation of the system, the Kalman gain and the system state uncertainty will gradually stabilize and fluctuate in a small range.

Fig. 3. The DOP and the number of received satellites in the study area on the day of the experiment.
IV. EXPERIMENT

A. Study Region

In this study, the performance of the proposed method is verified by using the taxi travel trajectory data in Shenzhen. The longitude range of the study area is 113.8 to 114.25, and the latitude range is 22.5 to 22.75. The study date is October 22, 2013. Fig. 5 shows the study area's administrative divisions, natural resources, land use, and urban road conditions.

Fig. 4. The variation range of the kalman gain and the uncertainty in the dynamic kalman filtering system.

Fig. 5. The geographic information of the study region.
B. Data Description

The data set records the taxi ID, current time, longitude, latitude, speed, and occupancy status. Table I shows some data.

<table>
<thead>
<tr>
<th>taxi ID</th>
<th>time</th>
<th>longitude</th>
<th>latitude</th>
<th>speed (m/s)</th>
<th>occupancy status</th>
</tr>
</thead>
<tbody>
<tr>
<td>22223</td>
<td>00:03:39</td>
<td>114.167732</td>
<td>22.562550</td>
<td>2.56</td>
<td>1</td>
</tr>
<tr>
<td>22223</td>
<td>00:03:54</td>
<td>114.168999</td>
<td>22.562550</td>
<td>11.94</td>
<td>1</td>
</tr>
<tr>
<td>22223</td>
<td>00:04:09</td>
<td>114.170998</td>
<td>22.562550</td>
<td>13.06</td>
<td>1</td>
</tr>
<tr>
<td>22223</td>
<td>00:04:23</td>
<td>114.172897</td>
<td>22.562599</td>
<td>15.11</td>
<td>1</td>
</tr>
<tr>
<td>22223</td>
<td>00:04:39</td>
<td>114.175232</td>
<td>22.561701</td>
<td>18.33</td>
<td>1</td>
</tr>
</tbody>
</table>

The field taxi ID is used to distinguish between different taxis and has no practical meaning. The time, longitude, latitude, and speed fields record an equipment's current time, position, and speed. The field occupancy status records whether the taxi has an ongoing order, with 1 meaning a passenger and 0 meaning no passenger. The data set contains 46927855 records of 14728 taxis. To display the spatial distribution of trajectory points in the study area, evenly divide the study area into 100 by 100 grids, count the number of records in each grid, and draw Fig. 6.

C. Data Analysis and Visualization

The entire dataset contains a large number of trajectories, which cannot be displayed in one picture, so just one taxi's trajectories are drawn. Fig. 7 shows 21 orders' trajectories for one taxi in a whole day. Due to trajectories having different lengths and too many trajectories are easy to overlap, it is unsuitable to be displayed in a figure with a constant scale. So, Fig. 7 is divided into four subfigures with different view ranges and scales. In Fig. 7, the solid line represents the trajectory that has not been optimized, and the dotted line represents the trajectory that the Kalman filter has optimized. Fig. 7(a), with the largest scale, shows some orders with long travel trajectories and long service duration, most of which occur late at night. Fig. 7(d), with the smallest scale, can more clearly compare the trajectories before and after optimization, which shows some orders in the city center.

D. Results

In order to more intuitively compare the accuracy of the trajectory before and after optimization, the deviation between the trajectory point and the nearest road is calculated, and accumulate the deviations by each order. The results show that the average deviation between trajectory points and urban roads is reduced by 33.6%, which indicates that the trajectory data optimization method based on Doppler velocimetry proposed in this study can reduce the deviation between trajectories and urban road networks and improve the quality of trajectory data. Notably, the deviation between trajectory points and urban roads is related to the error level but cannot directly represent it. For example, in the trajectory of the 19th order in Fig. 7(d), there is a section of the trajectory with a significant deviation. Still, the deviation value of the entire order is insignificant due to the lack of positioning points in this trajectory section. Sometimes the optimized trajectory cannot reduce the deviation, such as the 7th order in Fig. 7(c). At the overall level, the proposed optimization method can be considered efficient.
V. CONCLUSIONS

With the development of the mobile Internet, the number of terminal devices with positioning capability has increased rapidly. However, due to cost and energy consumption, the positioning capabilities of these devices have significant differences. This paper establishes the system state equation of travel trajectory using Doppler velocity measurement and pseudo-range positioning data. By correlating the measured random error in the system state with the DOP of satellite positioning, a dynamic Kalman filter system is established to
optimize the trajectory data. The results show that the optimized trajectory data can better fit the urban road and can be applied to data analysis in transportation studies.

ACKNOWLEDGMENT

This research was funded by the National Natural Science Foundation of China No. 61963011, Guangxi Science and Technology Major Project No. AA19254016, Project of Natural Science Youth Foundation of Guangxi Province No. 2020JJB170049.

REFERENCES


